



Modeling speech intelligibility in adverse conditions

Dau, Torsten

Published in:
The Listening Talker

Publication date:
2012

Document Version
Early version, also known as pre-print

[Link back to DTU Orbit](#)

Citation (APA):
Dau, T. (2012). Modeling speech intelligibility in adverse conditions. In *The Listening Talker: Proceedings* (pp. 6-6)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Torsten Dau

Danish Technical University

Modeling speech intelligibility in adverse conditions

In everyday life, the speech we listen to is often mixed with many other sound sources as well as reverberation. In such situations, people with normal hearing are able to almost effortlessly segregate a single voice out of the background. In contrast, hearing-impaired people have great difficulty understanding speech when more than one person is talking, even when reduced audibility has been fully compensated for by a hearing aid. The reasons for these difficulties are not well understood. This presentation highlights recent concepts of the monaural and binaural signal processing strategies employed by the normal as well as impaired auditory system. Jørgensen and Dau [(2011). *J. Acoust. Soc. Am.* 130, 1475-1487] proposed the speech-based envelope power spectrum model (sEPSM) in an attempt to overcome the limitations of the classical speech transmission index (STI) and speech intelligibility index (SII) in conditions with nonlinearly processed speech. Instead of considering the reduction of the temporal modulation energy as the intelligibility metric, as assumed in the STI, the sEPSM applies the signal-to-noise ratio in the envelope domain (SNR_{env}). This metric was shown to be the key for predicting the intelligibility of reverberant speech as well as noisy speech processed by spectral subtraction. However, the sEPSM cannot account for speech subjected to phase jitter, a condition in which the spectral structure of speech is destroyed, while the broadband temporal envelope is kept largely intact. In contrast, the effects of this distortion can be predicted successfully by the spectro-temporal modulation index (STMI) [Elhilali et al., (2003). *Speech Commun.* 41, 331-348], which assumes an explicit analysis of the spectral modulation energy. However, since the STMI applies the same decision metric as the STI, it fails to account for spectral subtraction. The results from the different modeling approaches suggest that the SNR_{env} might be a key decision metric while some explicit across-frequency pre-processing seems crucial to extract relevant speech features in some conditions.